A Hedonic Price Index of the House Price Statistics

Hannu Lehtimäki, Markku Verkkoniemi
Statistics Finland

1. INTRODUCTION

The statistics on house prices describe the unencumbered prices per square metre and price movements of old housing corporation flats in deals brokered by real estate agents.

The statistics comprise some 30 per cent of all housing deals concluded. The coverage of the municipalities included in statistics varies between 25-50 per cent. The regional coverage of the price statistics is greatest for the towns of Northern Finland in which the statistics include close on 40 per cent of the deals on old housing corporation flats.

A new method has been introduced into the measuring of the price development of dwellings. The aim of the change was to take into account better than before, not only the price data but also the effects of quality differences when measuring price development. The new index, which takes into account the quality differences, is a so-called hedonic price index. This index uses the price and quality data obtained from concluded housing deals when accounting for the price developments of dwellings. The new method is capable of solving the problem of the quality difference in the house price index clearly better than the earlier methods.

The house price index aims at describing how much more or less on average one has to pay for a dwelling of equal quality in the observation period in relation to the price of the base period. If the dwellings sold in each time period were exactly the same as to their characteristics, the measuring of price movements would be easy. For instance, the average price per square metre for dwellings could be used for measuring the price development. Since the population of dwellings sold in each particular quarter is in fact different, the changes in the average prices per square metre do not describe the pure price development, but the prices are also affected by the other characteristics of the sold dwellings. The quality characteristics of dwellings sold in some quarters may be better than those of the dwellings sold in another quarter. It is not expedient to compare the prices of dwellings of heterogeneous quality.

The hedonic price index is based on the idea that the prices of dwellings are defined on the market on the basis of their characteristics. By means of the prices and characteristics of the dwellings sold on the market it is possible to estimate a so-called hedonic price model which combines the prices and the characteristics. The model used is a regression model in which the price per square metre is explained by different quality attributes of the dwellings. By incorporating dummy variables that relate to the quarters into the price model it is possible to estimate the price development of dwellings, as the effect of the other attributes has been standardised. The coefficients of

the time dummy variables describe the relationship between dwellings of identical characteristics in different quarters.

Both the old and the new method give fairly similar estimates on the price development for the whole country, Greater Helsinki, the rest of Finland and Helsinki. As to smaller towns and areas, the differences between the old and the new index are in some cases fairly big. The new method is capable of taking into account the changes in the regional distribution of sold dwellings within the sales areas, which is the biggest improvement on the previous method. Earlier, if an exceptionally great proportion of the dwellings in some city were sold in the best residential areas of the said city, the price index was biased upward. Previously, the regional distribution of the sold dwellings could thus affect the point figure of the index directly.

The condition of the dwellings is another important characteristic the impact of which has not been taken into account in the price index calculation. The proportion of dwellings that are in good condition varies from one quarter to another even drastically and this has a clear effect on the price per square metre. The new model allows for the impact of the condition of the dwelling to be taken into account.

2. CALCULATION METHOD

The data known about each dwelling sold include the purchase price, the location with the accuracy of the postal code, age, floor area, type of building, number of rooms and date of sale. As of the beginning of 1994, the estimated condition is also known of nearly all dwellings. For some dwellings, the data also include the number of floors in the building, the ownership of the building lot and the rent. However, the data cannot be used in modelling as they are missing for too many dwellings.

2.1 Model

The variation observed in the prices per square metre are thought to derive from three main factors: the (difference between the) individual characteristics of the dwelling, the effect of the location of the dwelling and the residential area, and the period of sale. The model used can be expressed with the following equation:

$$(1)\quad LogP_{iat} = \alpha + \sum_{k=1}^{p} \beta_k x_{iak} + \sum_{a=2}^{Po} \mu_a A_{ai} + \sum_{t=2}^{J} \lambda_t T_{ti} + \varepsilon_{iat}$$

The dependent variable is the price per square metre expressed in terms of a logarithm log P_{iat}, where the subscript refers to the number of the observation, a to the area to which the observation i applies and t to the period of sale. x_{iak} refers to the value of the characteristic k in the observation i, which is located in the area a. p is the number of variables relating to the characteristics of the dwelling. A_{ia} s are dummy variables which take the value 1, if the dwelling i has been sold in the area in question, i.e. a. P_o is the number of areas. The dummy variables number one less than the areas in the area division used. The most common area division is that of postal

areas. T_{it} is a dummy variable that takes the value 1, if the observation i is sold in the time period t. J is the number of time periods. The time dummies number one less than the time periods, i.e. quarters of the year in practice. α,β_k,μ_a and λ_t are parameters. The model has a total of $p+(P_o-1)+(J-1)$ regressors proper.

Model equation (1) may be expressed in a form which gives the same OLS-estimator for the parameter vector as equation (1):

$$(2)\ LogP_{iat} = LogP_{\bar{a}} + \sum_{k=1}^{p} \beta_k (x_{iak} - x_{\bar{a}k}) + \sum_{t=2}^{J} \lambda_t (T_{it} - T_{\bar{a}t}) + (\varepsilon_{iat} - \varepsilon_{\bar{a}})$$

Model equation (2) is more handy than equation (1), since the area dummies may be replaced by area means. The models are estimated in this form.

2.2 Regressors

There are 10 variables relating to the individual characteristics of a dwelling in the model used:

0 = The mean of the logarithm of the price per square metre by postal code

1 and 2 = The floor area and the square root of the floor area. The square root is included because the correlation between the floor area and the logarithm of the price per square metre is not linear.

3 and 4 = The age and the square root of the age.

5 = The dummy variable for a block of flats: takes the value 1 if the dwelling sold is a block of flats.

6 = Number of rooms.

7, 8 and 9 State of repair good, satisfactory and unknown. Dwellings are classified into three categories according to their state of repair: good, bad and satisfactory. There are, however, a considerable number of dwellings for which the code is erroneous or the state of repair is unknown. An individual dummy variable is given to these dwellings. The dwellings in satisfactory repair form a comparison group.

In addition, there are 7 time dummies that relate to the quarters of the year.

2.3 Area division used

Postal areas are used as the area division.

Classification by area and type of building:

The house price index is calculated with a very dense area division.
The area indices are obtained by choosing the observations of the area in question and by forming a separate model of the observations of that area.

2.4 Calculation of the index

The price trend estimate is based on the coefficients of the time dummies in the model. As the dependent variable is expressed in terms of the logarithm, the coefficient estimate of the time dummy gives the change in the price level on the base period in log per cents. The log per cents are converted to ordinary percentages by means of a conversion formula. The index number is thus calculated in relation to the base period.

The indices are obtained from the model that conforms to the hedonic regression analysis. The logarithm of the price per square metre is the dependent variable. The regressors are the average price by postal area, floor area, age, square root of the floor area and age, the dummy variables formed by the type and repair of the dwelling and the dummy variables describing each quarter of the year. The regressors have been orthogonalised. To give the model stability, data for two years is always included in calculating the indices. As new quarters enter the calculation, older quarters are dropped. The first quarter of 1994 is naturally always included, because it is the base time period. The indices are calculated by type of dwellings and by area.

The index proper is obtained from the formula

$$IND_{yyq}=IND_{952}*EXP(\lambda_{yyq}-\lambda_{952}),$$

where IND_{yyq} is the index in question, IND_{952}=the index for the second quarter of 1995, λ_{yyq}= the coefficient of the DUMMY variable in question and λ_{952}=the coefficient of the DUMMY variable for the second quarter of 1995, which is the last time when the index was produced by the old way.

3. ADP SOLUTION

The PC application has been constructed using SAS/AF®. It is divided into three parts: the data input and checking, the calculation of indices for the quarter in question and their updating in the index file, and the output of the distribution data for the indices and the dwelling prices.

Real estate agents supply Statistics Finland with data on housing deals concluded during the quarter in question in both electronic form and on paper forms. These total over 5 000 during one quarter for the whole of Finland. For the data obtained in electronic form there are reading programs, whereas for the manual feeding a separate template with error-detection has been constructed with the help of PROC FSEDIT. Before the actual index run several checks are made on the material. The size class of the price per square metre is checked and any exorbitant or give-away prices are removed. The missing codes for the number of rooms and floor areas are listed and corrected. An attempt is made to estimate the missing floor area with the help of the number of rooms, and correspondingly the number of rooms with the help of the floor area. If both data are missing, the observation is left out of the calculation. Erroneous postal codes are listed and cor-

rected. The numbers of deals by postal area are listed for any missing observations. SAS/INSIGHT® software can be used for these checks.
In the index run the data for the quarter in question are first combined with the data for the previous quarters so that the material consists of data for two years. This procedure makes the parameters to be estimated in the index run more stable. The price indices are obtained by PROC REG from the coefficients of the time dummies by multiplying by a certain constant. In addition, various key indicators relating to the price breakdown of the dwellings are calculated in the index run. These include, for example, the average price per square metre, median, 10 per cent and 90 per cent deciles and number of deals. Other SAS® procedures used are PROC MEANS and PROC UNIVARIATE.

The output tables are constructed from the SAS® file using Microsoft® Excel. The tables comprise indices by dwelling type for 36 different area divisions. In addition, it is possible to select the house price index for different base years as well as indices in relation to the cost-of-living index and the building cost index.

There are two permanent SAS®* files in the application. First, the file for real estate agents which enables us to make different ad hoc compilations from the material besides giving standard results. Second, the index file containing indices by area and dwelling type, average prices per square metre, number of deals and change percentages on the previous quarter and year.

* SAS/AF, SAS/INSIGHT and SAS are registered trademarks of SAS Institute Inc., Cary, NC, USA.